

09/720934

SH3D1A Domain Structure and Homologies - Human vs Xenopus (Determined using GCG programs, BLAST, FASTA)

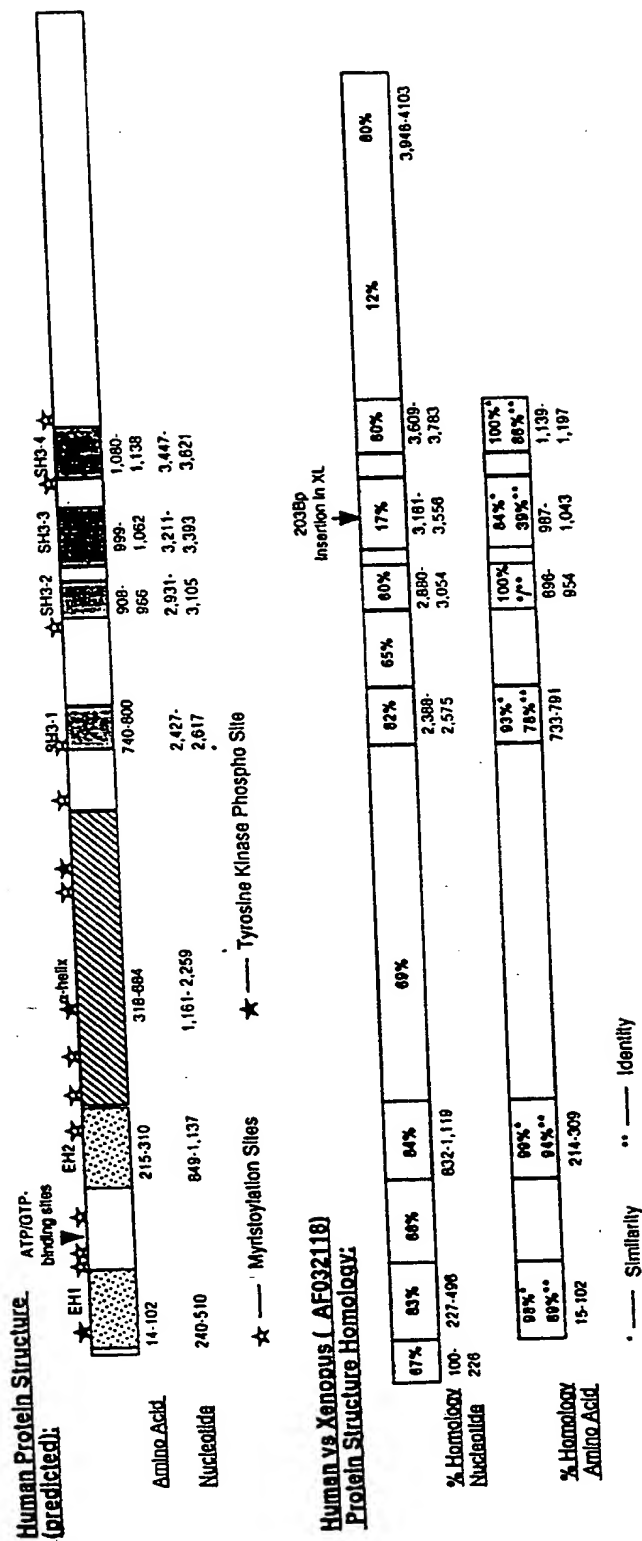
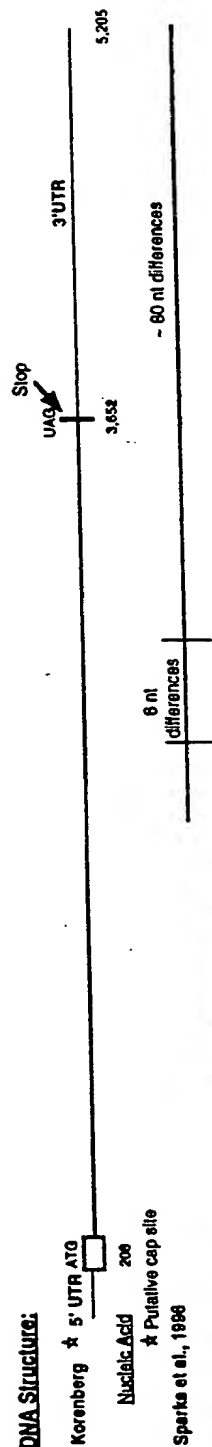
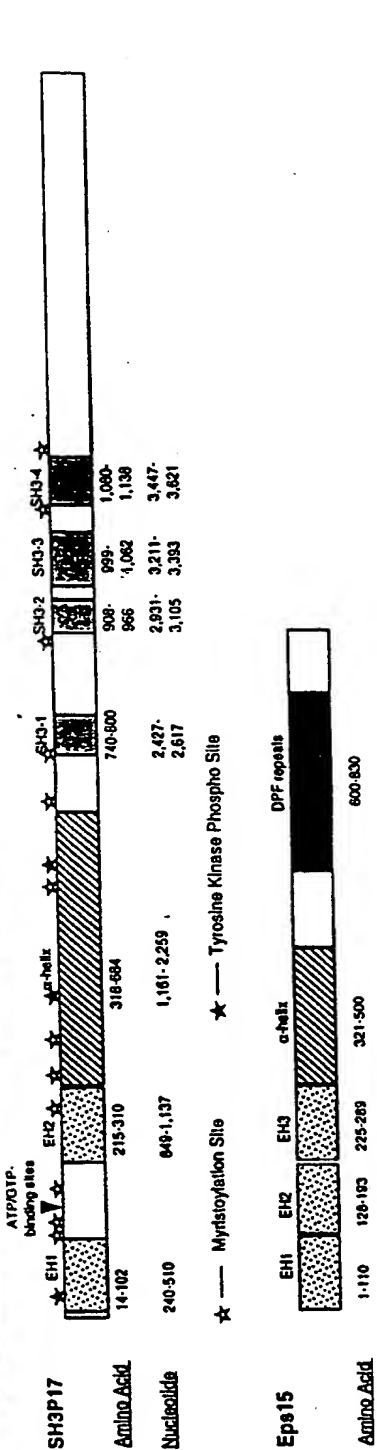


Figure 1

cdNA Structure:



EH Domain Comparison of Human SH3P17 and Mouse Eps15



Scale for Above Ideograms:

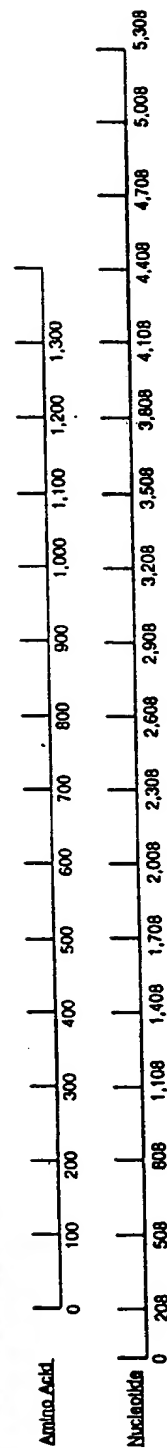


Figure 2

Region of Chromosome 21 Responsible for Megakaryocytic Abnormalities

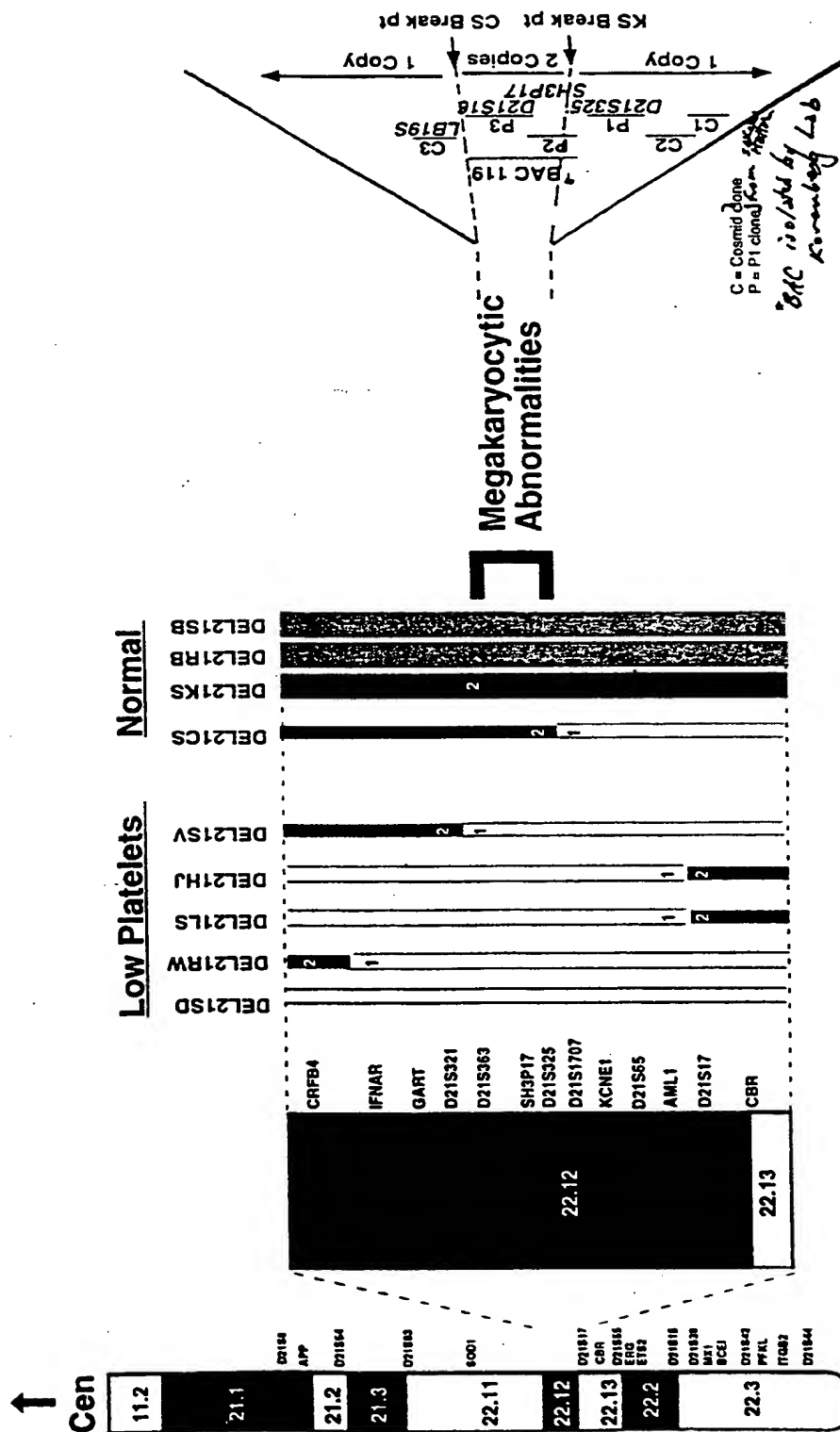


Figure 3

SH3D1A

1 CAAAAGAATT CCGGGTACGG CGGCTCGOGA GGAAGAATCC CGAGCGGGCT
51 CCGGGACGGA CAGAGAGGCG GCGGGGATG GTGTGOGGGG CTGCGGCTCC
101 TGCGTCCCTC CCAGCGGGCG GTGAGCGGCA CTGATTTGTC CCTGGGGGCG
151 CAGCGGGAC CCGCCCGGAG ATGAGGCGTC GATTAGCAAG GTAAAAGTAA
201 CAGAACCATG GCTCAGTTTC CAACACCTTT TGGTGGCAGC CTGGATATCT
251 GGGCCATAAC TGTAGAGGAA AGACCGAAGC ATGATCAGCA GTTCATAGT
301 TTAAAGCCAA TATCTGGATT CATTACTGGT GATCAAGCTA GAACTTTTTT
351 TTTTCAATCT GGGTTACCTC AACCTGTTTT AGCACAGATA TGGGCACTAG
401 CTGACATGAA TAATGATGGA AGAATGGATC AAGTGGAGTT TTCCATAGCT
451 ATGAAACTTA TCAAACGAA GCTACAAGGA TATCAGCTAC CCTCTGCACT
501 TCCCCCTGTC ATGAAACAGC AACCAGTTGC TATTTCTAGC GCAACAGCAT
551 TTGGTATGGG AGGIATCGCC AGCATGCCAC CGCTTACAGC TGTGTCTCCA
601 GTGCCAATGG GATCCATTCC AGTTGTGGA ATGTCTCCAA CCTAGTATC
651 TTCTGTTCCT ACAGCAGCTG TGCCCCCCTT GGCTAAAGGG GCTCCCCCTG
701 TTATACAACC TCTGCTGCA TTTGCTCATC CTGCAGCCAC ATTGCCAAG
751 AGTTCTTCCT TTAGTAGATC TGGTCCAGGG TCACAACTAA AACTAAATT
801 ACAAAGGCA CAGTCATTIG ATGTGGCCAG TGTCCACCA GTGCCAGAT
851 GGGCTGTTCC TCAGTCATCA AGACTGAAAT ACAGGCAATT ATTCAATAGT
901 CATGACAAAA CTATGAGTGG AACTTAACA GTTCCCCAAG CAAGAACTAT
951 TCTTATGCAG TCAAGTTTAC CACAGGCTCA GCTGGCTTCA ATATGGAATC
1001 TTTCTGACAT TGATCAAGAT GGAAACTTA CAGCAGAGGA ATTATCCTG

Figure 4

1051 GCAATGCACC TCATTGATGT AGCTATGTCT GGCCAACCAC TGCCACCTGT
1101 CCTGCCTCCA GAATACATT CACCTTCTTT TAGAAGAGTT CGATCTGGCA
1151 GTGGTATATC TGTCATAAGC TCAACATCTG TAGATCAGAG GCTACCAGAG
1201 GAACCAGTTT TAGAAGATGA ACAACAACAA TTAGAAAAGA AATTACCTGT
1251 AACGTTTGAA GATAAGAAGC GGGAGAACTT TGAACGTGGC AACCTGGAAC
1301 TGGAGAAACG AAGGCAAGCT CTCCTGGAAC AGCAGCGCAA GGAGCAGGAG
1351 CGCCTGGCCC AGCTGGAGCG GGCGGAGCAG GAGAGGAAGG AGCGTGAGCG
1401 CCAGGAGCAA GAGCGCAAAA GACAACCTGA ACTGGAGAAG CAACTGGAAA
1451 AGCAGCGGGA CCTAGAACGG CAGAGAGAGG AGGAGAGGAG GAAAGAAATT
1501 GAGAGCGGAG AGGCTGCAAA ACGGAACTT GAAAGGCAAC GACAACCTGA
1551 GTGGGAACGG AATCGAAGC AAGAACTACT AAATCAAAGA AACAAAGAAC
1601 AAGAGGACAT AGTTGTACTG AAACCAAGA AAAAGACTTT GGAATTTGAA
1651 TTAGAAGCTC TAAATGATAA AAAGCATCAA CTAGAAGGGA AACTTCAAGA
1701 TATCAGATGT CGATTGACCA CCCAAAGGCA AGAAATTGAG AGCACAACA
1751 AATCTAGAGA GTTGAGAATT GCGAAATCA CCCATCTACA GCAACAATTA
1801 CAGGAATCTC AGCAAATGCT TGGAAGACTT ATTCCAGAAA AACAGATACT
1851 CAATGACCAA TTAAAACAAG TTCAGCAGAA CAGTTTGCAC AGAGATTAC
1901 TTGTTACTT TAAAGAGCC TTAGAAGCAA AAGAACTAGC TOGGCAGCAC
1951 CTACGAGACC AACTGGATGA AGTGGAGAAA GAAACTAGAT CAAACTACA
2001 GGAGATTGAT ATTTTCAATA ATCAGCTGAA GGAACATAA GAAATACACA
2051 ATAAGCAACA ACTCCAGAAG CAAAAGTCCA TGGAGGCTGA ACGACTGAAA
2101 CAGAAAGAAC AAGAACGAAA GATCATAGAA TTAGAAAAAC AAAAAGAAGA
2151 AGCCCAAAGA CGAGCTCAGG AAAGGGACAA GCAGTGGCTG GAGCATGTGC
2201 AGCAGGAGGA CGAGCATCAG AGACCAAGAA AACTCCACGA AGAGGAAAAA
2251 CTGAAAAGGG AGGAGAGTGT CAAAAGAAG GATGGCGAGG AAAAAGGCAA

Figure 4

2301 ACAGGAAGCA CAAGACAAGC TGGGTGGCT TTTCATCAA CACCAAGAAC
2351 CAGCTAAGCC AGCTGTCCAG GCACCTGGT CCACTGCAGA AAAAGGTCCA
2401 CTTACCATTT CTGCACAGGA AAATGTAAA GTGGTGTATT ACGGGGCACT
2451 GTACCCCTTT GAATCCAGAA GCCATGATGA AATCACTATC CAGCCAGGAG
2501 ACATAGTCAT GGTGGATGAA AGCCAACTG GAGAACCCG CTGGCTTGGA
2551 GGAGAATTAA AAGGAAAGAC AGGTTGGTTC CCTGCAAACT ATGCAGAGAA
2601 AATCCCAGAA AATGAGGTTT CCGTCCAGT GAAACCAGTG ACTGATTCAA
2651 CATCTGCCCC TGCCCCAAA CTGGCTTGC GTGAGACCCC CGCCCCTTTG
2701 GCAGTAACT CTTGAGAGCC CTCCAGACC CCTAATAACT GGGCCGACTT
2751 CAGCTCCAGG TGGCCCCACCA GCACGAATGA GAAACCAGAA ACGGATAACT
2801 GGGATGCATG GGCAGCCAG CCGTCTCTCA CCGTCCAAG TGCCGGCCAG
2851 TTAAGGCAGA GGTCCGCTT TACTCCAGCC ACGGCCACTG GCTCTCCCC
2901 GTCTCTGTG CTAGGCCAGG GTGAAAAGGT GGAGGGGCTA CAAGCTCAAG
2951 CCGTATATCC TTGGAGAGCC AAAAAGACA ACCACTTAA TTTTAACAA
3001 AATGATGTCA TCACCGTCTT GGAACAGCA GACATGTGGT GGTTTGGAGA
3051 AGTTCAAGGT CAGAAGGTT GTTCCCCAA GTCTTACGTG AAATCATTT
3101 CAGGGCCCAT AAGGAAGTCT ACAAGCATGG ATTCTGGTTC TTCAGAGAGT
3151 CCTGCTAGTC TAAAGGAGT AGCCTCTCCA GCAGCCAAGC CGGTGTTTC
3201 GGGAGAAGAA ATTGCCCAGG TTATTGCTC ATACACCGCC ACCGGCCCCG
3251 AGCAGCTCAC TCTGCCCCCT GTTCAGCTGA TTTTGATCG AAAAAGAAC
3301 CCAGGTGGAT GGTGGGAAGG AGAGCTGCAA GCAGTGGA AAAAGGCCA
3351 GATAGGCTGG TTCCAGCTA ATTATGTAAA GCTCTAAGC CCTGGGACGA
3401 GCAAAATCAC TCCAACAGAG CCACCTAAGT CAACAGCATT AGCGGCAGTG
3451 TGCCAGGTGA TTGGGATGTA CGACTACACC GGCAGAAATG ACGATGAGCT

Figure 4

3501 GGCCTTCAAC AAGGGCCAGA TCATCAAGT CCTCAACAAG GAGGACCTG
3551 ACTGGTGGAA AGGAGAAGTC AATGGACAAG TGGGGCTCTT CCCATCCAAT
3601 TATGTGAAGC TGACCACAGA CATGGACCCA AGCCAGCAAT GAATCATATG
3651 TTGTCCATCC CCCCCCAGG CTTGAAAGTC CTCAAAGAGA CCCACTATCC
3701 CATATCACTG CCCAGAGGGA TGATGGGAGA TGCAGCCTTG ATCATGTGAC
3751 TTCCAGCATG ATCACCTACT GCCTTCIGAG TAGAAGAACT CACTGCAGAG
3801 CAGTTTACCT CATTTTACCT TAGTTGCATG TGATCCCAAT GTTTGAGTTA
3851 TTACTTGCAG AGATAGGAGC AAAAATTACA AAAACACACA GGGTAGTGGG
3901 TCCTTTGTG GCTTTCCTAG TTAATCAAAT TGACTTTCCO CCACCTTTGC
3951 ACAGGTGCTT TCAATAGTTT TAAAATTATT TTAAATATA TATTTTAGCT
4001 TTTTAATAAA CAAAATAAAT AAATGACTTC TTTGCTATTT TGGTTTTCGA
4051 AAAAGACCCA CTATCAAGGA ATGCTGCATG TGCTATTAAA AATGTTCOCA
4101 AATGTCCATA AATCTGAGAC TIGATGTATT TTTTCATTTT GTCCAGTGT
4151 ACCAACTAAA TTGCTGCAGT TTGGGGCTTT TCCCCCTAC CATAGAAGTG
4201 CAGAGGAGTT CAGTATCTCT GTTTTAAAGA CGTATAGAAT GAGCCCAATT
4251 AAAGCGAAGG TGATTGTGCT TGTGTGTGTG TATCAGCTGT ACCTTGTGTA
4301 GCATGTAAATA CATCTGTAC ATAAGAAATT AGTCTTTCC ATGGCAAAGC
4351 TATTACCTTG TACGATGCTC TAATCATATT GCATTTAATT TTATTTTGCA
4401 ACAGTGACCT TGTAGCCACA TGAGAAAGCA CTCTGTGTTT TTGTTOGGTC
4451 TCAGATTTAT CTGGTTGAGT TGGTGTGTTG TTTGGGGTTT TTAATTTTGC
4501 GTGTTTGCAT AGCATAAAAT CAGTAGACAA CACCACTGAG GTCGTTACGA
4551 TCAACGATAT CCACAGTCTC TTTTAGTCT CTGTACATG AAGTTTTATT
4601 CCAGTTACTT TTCATGGAAT GACCTATTTT GAACAAGTAA TTTTCTTGAC
4651 AAGAAAGAAT GTATAGAAGT CTCCCTGCAA TTAATTTCCA ATGTTTACAT
4701 TTTTAACTA GGACTGTGGA ATTTCTACAG ATTAAATATGA AATGGAGCTC

Figure 4

4751 ATGGTCCGTT TGTTGTGTTAG ATATGCTGTA GCTGAAGCCC TGTTTGTCTT
4801 TTAAACACTA GTTGGGAAGCT CTCAATAAAA ATGCTGTCTG CTCACAGCAC
4851 AGAAAATGGG GCAGGGGGAG CCTCAAGCAC AATCTAGCTG TCTCTCTAAA
4901 GACTCTGTAA TGCTCAATCC CCTTGGCTTC TCCCGGGGCT GTCGGGAGGC
4951 TGTCCTGGTG GTGGGTGAGA GGTCTTTTTC CTTTCAAATG GTGCAGAGAG
5001 AGAGGACCTT TCTCTCTTGT TCAGTTGCAA TTCAGTATTT TCACGGATAT
5051 GAATGTAAAA TATATAAATA TATAAACCTG AGGATTTAAC AAATGTAAAA
5101 CAACCTTTTG AATTAGTTCC GAGTATAGAT AATTAAATTT TTAAAACAAA
5151 AGTAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGTGGAC GCGGCGCGG

Figure 4

SH3D1A Translated Protein Sequence:

1 MAQFPITFFGG SLDIWAITVE ERAKHDQQFH SLKPISGFIT GDQARNFTFQ
51 SGLPQPVLAQ IVALADMND GRMDQVEFSI AMKLIKLLKQ GYQLPSALPP
101 VMKQOPVAIS SAPAFGMGGI ASMPPLTAVA FVMGSIPIVV GMSPTLVSSV
151 PTAAVPPLAN GAPPVIOPLP AFAHPAATLP KSSSFSSRSGP GSQLNKLOK
201 AQSFVAVSP PVAEWAHPQS SRLKYROLFN SHKIMSGHL TGFQARTILM
251 QSSLPQAQLA SIWNLSDDIQ DGKLTAEEFI LAMHLIDVAM SGQPLPFVLP
301 PEYTPPSFRR VRSGSGISVI SSTSVDRQLP EEPVLEDEQQ QLEKKLPVTF
351 EDKKRENFER GNLELEKRRQ ALLEQORKEQ ERLAQLERAQ QERKERERQE
401 QERKQLELE KQLEKQRELE RQREERREKE TERREAARE LERQOLEWE
451 RNRQELLNQ RNKEQEDIVV LKAKKILEF ELEALNDKKH QLEGKLODIR
501 CRLTQROEI ESTNKSREL RIAETHLQQQ LQESQMLGR LIPEKQILND
551 QLKQVQNSL HRDSLVTIKR ALEAKELARQ HLRDQLEVE KETRSLQEI
601 DIFNNQKEL REIHNKQQLQ KQKSMEARL KQKEQERKII ELEKQKEEAQ

Figure 5

651 RRAQERDKQW LEHVQOEDEH QRPRKLHEEE KKKREESVKK KDGEEKGKQE
701 AQDKLGRLFH QHQEPAPAV QAPWSTAERG PLTISAQENV KVVYYRALYP
751 FESRSHDEIT IQPGDIVMVD ESQTGEPGWL GGELKGTGW FPANYAEKIP
801 ENEVPAPVKP VTDSTSAPAP KLALRETPAP LAVTSSEPST TENWADFSS
851 TWPTSTNEKP ETDNDAAWA QPSLTVPSAG QLRQSAFTP ATATGSSPSP
901 VLGQGEKVEG LQACALYPWR AKKDNHLNEN KNDVITVLEQ QDMWWFGEVQ
951 GQKGWFPKSY VKLISGPIRK STSMDSGSSE SPASLKRVAS PAAKPVVSGE
1001 EIAQVIASYT ATGPEQ/ILA PQQLILIRKK NPGGWEGEL QARGKKRQIG
1051 WFPANYVKLL SPGTSKITPT EPPKSTALAA VQVIGMYDY TAQNDDELAF
1101 NKGQIINVLN KEDPDWWKGE VNGQVGLFPS NYVKLTIDMD PSQ

Figure 5

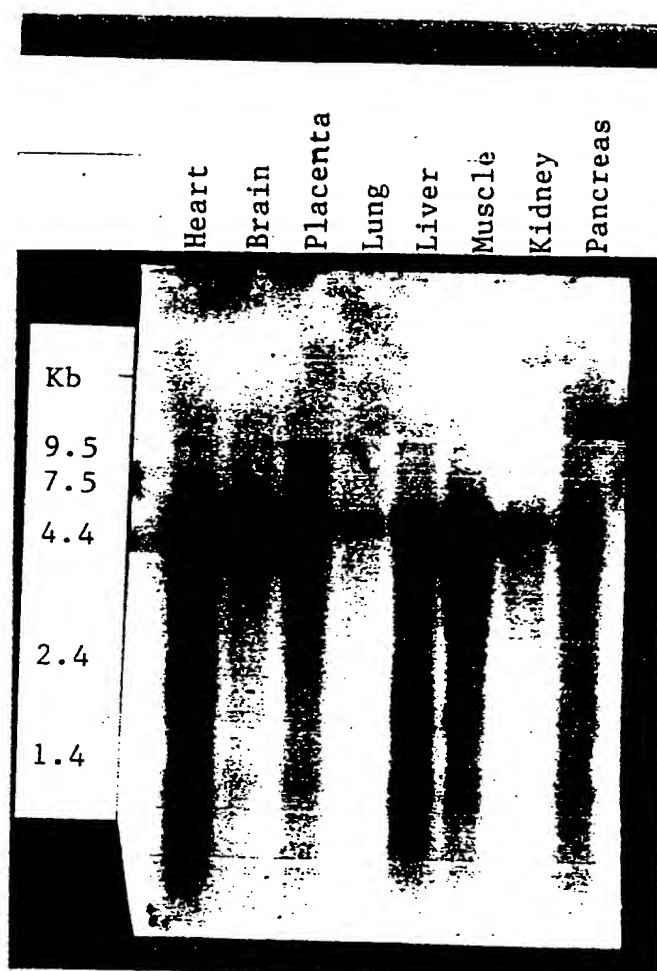


Figure 6

Summary of cDNAs Isolated

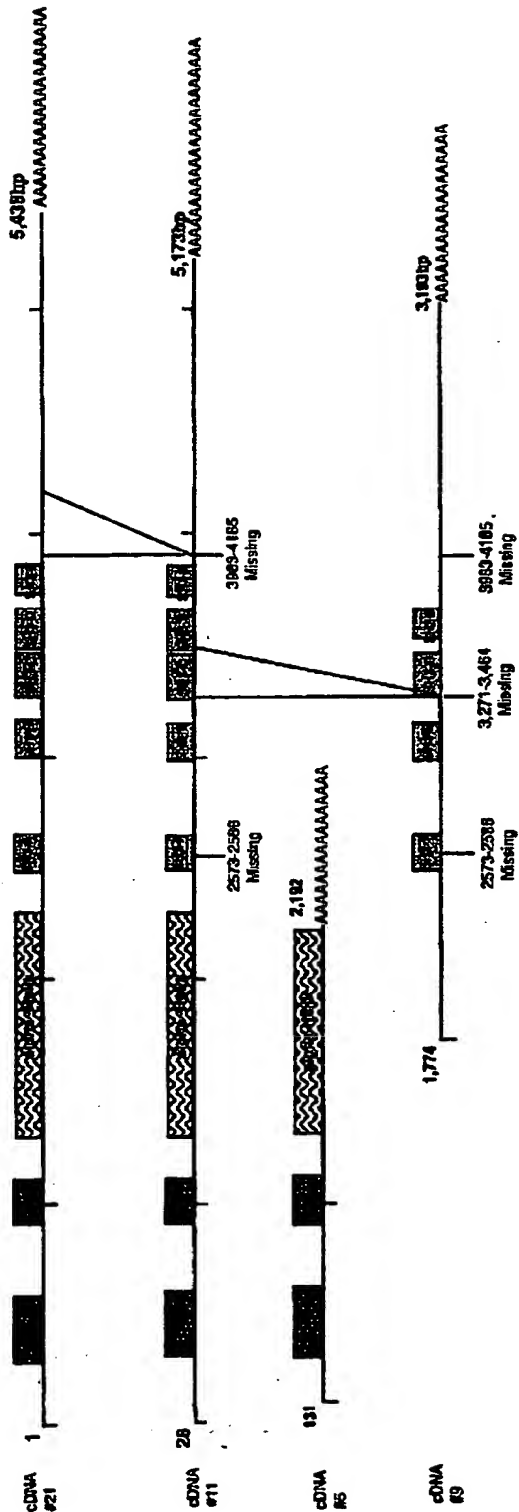


Figure 7

1 GCACGAGAGG GAGCGAAGGA GG TAGAGAAG AGTGGAGGCG CCAGGGGAGG
51 GAGCGTAGCT TGGTTGCTCC GTAGTACGGC GGCTCGCGAG GAAGAATCCC
101 GAGCGGGCTC CGGGACGGAC AGAGAGGCGG GCGGGGATGG TGTGCGGGGC
151 TGCGGCTCCT GCGTCCCTCC CAGCGGCGCG TGAGCGGCAC TGATTTGTCC
201 CTGGGGCGGC AGCGCGGACC CGCCCGGAGA TGAGGCGTCG ATTAGCAAGG
251 TAAAAGTAAC AGAACCATGG CTCAGTTTCC AACACCTTTT GGTGGCAGCC
301 TGGATATCTG GGCCATAACT GTAGAGGAAA GAGCGAAGCA TGATCAGCAG
351 TTCCATAGTT TAAAGCCAAT ATCTGGATTG ATTACTGGTG ATCAAGCTAG
401 AAACCTTTTTT TTTCAATCTG GGTTACCTCA ACCTGTTTTA GCACAGATAT
451 GGGCACTAGC TGACATGAAT AATGATGGAA GAATGGATCA AGTGGAGTTT
501 TCCATAGCTA TGAAACTTAT CAAACTGAAG CTACAAGGAT ATCAGCTACC
551 CTCTGCACTT CCCCCTGTCA TGAAACAGCA ACCAGTTGCT ATTTCTAGCG
601 CACCAGCATT TGGTATGGGA GGTATCGCCA GCATGCCACC GCTTACAGCT
651 GTTGCTCCAG TGCCAATGGG ATCCATTCCA GTTGTGGAA TGTCTCCAAC
701 CCTAGTATCT TCTGTTCCA CAGCAGCTGT GCCCCCCTG GCTAACGGGG
751 CTCCCCCTGT TATACAACCT CTGCCTGCAT TTGCTCATCC TGCAGCCACA
801 TTGCCAAAGA GTTCTTCCTT TAGTAGATCT GGTCCAGGGT CACAATAAA
851 CACTAAATTA CAAAAGGCAC AGTCATTTGA TGTGGCCAGT GTCCCACCAG
901 TGGCAGAGTG GGCTGTTCTT CAGTCATCAA GACTGAAATA CAGGCAATTA
951 TTCAATAGTC ATGACAAAAC TATGAGTGGA CACTTAACAG GTCCCCAAGC
1001 AAGAACTATT CTTATGCAGT CAAGTTTACC ACAGGCTCAG CTGGCTTCAA
1051 TATGGAATCT TTCTGACATT GATCAAGATG GAAAACCTAC AGCAGAGGAA
1101 TTTATCCTGG CAATGCACCT CATTGATGTA GCTATGTCTG GCCAACCCT
1151 GCCACCTGTC CTGCCTCCAG AATACATTCC ACCTTCTTTT AGAAGAGTTC
1201 GATCTGGCAG TGGTATATCT GTCATAAGCT CAACATCTGT AGATCAGAGG
1251 CTACCAGAGG AACCAGTTTT AGAAGATGAA CAACAACAAT TAGAAAAGAA
1301 ATTACCTGTA ACGTTTGAAG ATAAGAAGCG GGAGAACTTT GAACGTGGCA
1351 ACCTGGAACCT GGAGAAACGA AGGCAAGCTC TCCTGGAACA GCAGCGCAAG
1401 GAGCAGGAGC GCCTGGCCCA GCTGGAGCGG GCGGAGCAGG AGAGGAAGGA
1451 GCGTGAGCGC CAGGAGCAAG AGCGCAAAAG ACAACTGGAA CTGGAGAAGC
1501 AACTGGAAAA GCAGCGGGAG CTAGAACGGC AGAGAGAGGA GGAGAGGAGG
1551 AAAGAAATTG AGAGGCGAGA GGCTGCAAAA CGGGAACCTG AAAGGCAACG
1601 ACAACTTGAG TGGGAACGGA ATCGAAGGCA AGAACTACTA AATCAAAGAA
1651 ACAAGAACA AGAGGACATA GTTGTACTGA AAGCAAAGAA AAAGACTTTG
1701 GAATTTGAAT TAGAAGCTCT AAATGATAAA AAGCATCAAC TAGAAGGGAA
1751 ACTTCAAGAT ATCAGATGTC GATTGACCAC CCAAAGGCAA GAAATTGAGA
1801 GCACAAACAA ATCTAGAGAG TTGAGAATTG CCGAAATCAC CCATCTACAG
1851 CAACAATTAC AGGAATCTCA GCAAATGCTT GGAAGACTTA TTCCAGAAAA
1901 ACAGATACTC AATGACCAAT TAAAACAAGT TCAGCAGAAC AGTTTGCACA
1951 GAGATTCACT TGTTACACTT AAAAGAGCCT TAGAAGCAAA AGAACTAGCT
2001 CGGCAGCACC TACGAGACCA ACTGGATGAA GTGGAGAAAAG AAAGTAGATC
2051 AAAACTACAG GAGATTGATA TTTTCAATAA TCAGCTGAAG GAACTAAGAG
2101 AAATACACAA TAAGCAACAA CTCCAGAAGC AAAAGTCCAT GGAGGCTGAA

Figure 8

2151 CGACTGAAAC AGAAAGAACA AGAACGAAAG ATCATAGAAT TAGAAAAACA
2201 AAAAGAAGAA GCCCAAAGAC GAGCTCAGGA AAGGGACAAG CAGTGGCTGG
2251 AGCATGTGCA GCAGGAGGAC GAGCATCAGA GACCAAGAAA ACTCCACGAA
2301 GAGGAAAAAC TGAAAAGGGA GGAGAGTGTC AAAAAGAAGG ATGGCGAGGA
2351 AAAAGGCAAA CAGGAAGCAC AAGACAAGCT GGGTCGGCTT TTCCATCAAC
2401 ACCAAGAACC AGCTAAGCCA GCTGTCCAGG CACCCTGGTC CACTGCAGAA
2451 AAAGGTCCAC TTACCATTTC TGCACAGGAA AATGTAAAAG TGGTGTATTA
2501 CCGGGCACTG TACCCCTTTG AATCCAGAAG CCATGATGAA ATCACTATCC
2551 AGCCAGGAGA CATAGTCATG GTTAAAGGGG AATGGGTGGA TGAAAGCCAA
2601 ACTGGAGAAC CCGGCTGGCT TGGAGGAGAA TTAAAAGGAA AGACAGGGTG
2651 GTTCCCTGCA AACTATGCAG AGAAAAATCCC AGAAAATGAG GTTCCCGCTC
2701 CAGTGAAACC AGTGACTGAT TCAACATCTG CCCCTGCCCC CAAACTGGCC
2751 TTGCGTGAGA CCCCCGCCCC TTTGGCAGTA ACCTCTTCAG AGCCCTCCAC
2801 GACCCCTAAT AACTGGGCCG ACTTCAGCTC CACGTGGCCC ACCAGCACGA
2851 ATGAGAAACC AGAAACGGAT AACTGGGATG CATGGGCAGC CCAGCCCTCT
2901 CTCACCGTTC CAAGTGCCGG CCAGTTAAGG CAGAGGTCCG CCTTTACTCC
2951 AGCCACGGCC ACTGGCTCCT CCCCCTCTCC TGTGCTAGGC CAGGGTGAAA
3001 AGGTGGAGGG GCTACAAGCT CAAGCCCTAT ATCCTTGGAG AGCCAAAAAA
3051 GACAACCACT TAAATTTTAA CAAAAATGAT GTCATCACCG TCCTGGAACA
3101 GCAAGACATG TGGTGGTTTG GAGAAGTTCA AGGTCAGAAG GGTTGGTTCC
3151 CCAAGTCTTA CGTGAAACTC ATTTACAGGGC CCATAAGGAA GTCTACAAGC
3201 ATGGATTCTG GTTCTTCAGA GAGTCCTGCT AGTCTAAAGC GAGTAGCCTC
3251 TCCAGCAGCC AAGCCGGTCG TTTCGGGAGA AGAATTTATT GCCATGTACA
3301 CTTACGAGAG TTCTGAGCAA GGAGATTAA CCTTTCAGCA AGGGGATGTG
3351 ATTTTGTTA CCAAGAAAGA TGGTACTGG TGGACAGGAA CAGTGGGCGA
3401 CAAGGCCGGA GTCTTCCCTT CTAACATATG GAGGCTTAAA GATTACAGAT
3451 GCTCTGGAAC TGCTGGGAAA ACAGGGAGTT TAGGAAAAAA ACCTGAAATT
3501 GCCCAGGTTA TTGCCTCATA CACCGCCACC GGCCCCGAGC AGCTCACTCT
3551 CGCCCCTGGT CAGCTGATTT TGATCCGAAA AAAGAACCCA GGTGGATGGT
3601 GGGAAGGAGA GCTGCAAGCA CGTGGGAAAA AGCGCCAGAT AGGCTGGTTC
3651 CCAGCTAATT ATGTAAAGCT TCTAAGCCCT GGGACGAGCA AAATCACTCC
3701 AACAGAGCCA CCTAAGTCAA CAGCATTAGC GGCAGTGTGC CAGGTGATTG
3751 GGATGTACGA CTACACCGCG CAGAATGACG ATGAGCTGGC CTTCAACAAG
3801 GGCCAGATCA TCAACGTCCT CAACAAGGAG GACCCTGACT GGTGGAAAGG
3851 AGAAGTCAAT GGACAAGTGG GGCTCTTCCC ATCCAATTAT GTGAAGCTGA
3901 CCACAGACAT GGACCCAAGC CAGCAATGAA TCATATGTTG TCCATCCCCC
3951 CCTCAGGCTT GAAAGTCCTC AAAGAGACCC ACTATCCCAT ATCACTGCCC
4001 AGAGGGATGA TGGGAGATGC AGCCTTGATC ATGTGACTTC CAGCATGATC
4051 ACCTACTGCC TTCTGAGTAG AAGAACTCAC TGCAGAGCAG TTTACCTCAT
4101 TTTACCTTAG TTGCATGTGA TCGCAATGTT TGAGTTATTA CTTGCAGAGA
4151 TAGGAGCAAA AATTACAAA ACACACAGGG TAGTGGGTCC TTTTGTGGCT
4201 TTCCTAGTTA CTCAAATTGA CTTTCCCCCA CCTTTGCACA GGTGCTTTCA
4251 ATAGTTTTAA AATTATTTT AAATATATAT TTTAGCTTTT TAATAAACAA
4301 AATAAATAAA TGACTTCTTT GCTATTTTGG TTTGCAAAA AGACCCACTA
4351 TCAAGGAATG CTGCATGTGC TATTAATAAT TGTTCCAAAT GTCCATAAAT

Figure 8

4401 CTGAGACTTG ATGTATTTTT TCATTTTGTC CAGTGTTACC AACTAAATTG
4451 TGCAGTTTGG GGCTTTTCCC CCTTACCATA GAAGTGCAGA GGAGTTCAGT
4501 ATCTCTGTTT TAAAGACGTA TAGAATGAGC CCAATTAAAG CGAAGGTGTT
4551 TGTGCTTGTT TGTGTGTATC AGCTGTACCT TGTTGAGCAT GTAATACATC
4601 CTGTACATAA GAAATTAGTT CTTTCCATGG CAAAGCTATT ACCTTGTACG
4651 ATGCTCTAAT CATATTGCAT TTAATTTTAT TTTGCACAGT GACCTTGTAG
4701 CCACATGAGA AAGCACTCTG TGTTTTTGTT CGGTCTCAGA TTTATCTGGT
4751 TGAGTTGGTG TTTTGTTTGG GGTTTTAAAT TTTGCGTGTT TGCATAGCAT
4801 AAAATCAGTA GACAACACCA CTGAGGTCGT TACGATCAAC GATATCCACA
4851 GTCTCTTTTT AGTCTCTGTT ACATGAAGTT TTATTCCAGT TACTTTTCAT
4901 GGAATGACCT ATTTTGAACA AGTAATTTTC TTGACAAGAA AGAATGTATA
4951 GAAGTCTCCC TGCAATTAAT TTCCAATGTT TACATTTTTT AACTAGACTG
5001 TGGAATTTCT ACAGATTAAT ATGAAATGGA GCTCATGGTC CGTTTGTGTG
5051 TTAGATATGC TGTAGCTGAA GCCCTGTTTG TCTTTTAAAC ACTAGTTGGA
5101 AGCTCTCAAT AAAAATGCCT GCTGCTCACA GCACAGAAAA TGGGGCAGGG
5151 GGAGCCTCAA GCACAATCTA GCTGTCCTCC TAAAGACTCT GTAATGCTCA
5201 CTCCCCTCGC GTTCTCCCGG CGCTGTCGGG AGGCTGTGCT GGTGGTCGTG
5251 TAGAGGTCCT TCTCCTTTCA CATGGTGCAG AGAGCGAGGA CCTCTCCTCC
5301 TCGTTCAGTT GCACTTCAGT ATTTTCACGG ATATGAATGT AAAATATATA
5351 AATATATAAA CCTGCGGCTT TAACAACTGT AATACAACCT TTTGAATTAG
5401 TTCCGTGTAT AGATAATTAA ATTCTTCATA CAAAAGTTAA AAAAAAAAAA
5451 AAAAAAAAAA

Figure 8

#21 translated protein sequence:

1 MAQFPTPFGG SLDIWAITVE ERAKHDQQFH SLKPISGFIT GDQARNFFFQ
51 SGLPQPVL AQ I WALADMNND GRMDQVEFSI AMKLIKLLKQ GYQLPSALPP
101 VMKQQPVAIS SAPAFGMGGI ASMPPLTAVA PVPMGSIPVV GMSPTLVSSV
151 PTAAVPPLAN GAPPVIQPLP AFAHPAATLP KSSSFSRSGP GSQLNTKLQK
201 AQSFDVASVP PVAEWAVPQS SRLKYRQLFN SHDKTMSGHL TGPQARTILM
251 QSSLPQAQLA SIWNLSIDIQ DGKLTAEFEI LAMHLIDVAM SGQPLPPVLP
301 PEYIPPSFRR VRSGSGISVI SSTVDQRLP EEPVLEDEQQ QLEKKLPVTF
351 EDKKRENFER GNLELEKRRQ ALLEQQRKEQ ERLAQLERAE QERKERERQE
401 QERKRQLELE KQLEKQRELE RQREEERRKE IERREAAKRE LERQRQLEWE
451 RNRRQELLNQ RNKEQEDIVV LKAKKKTLEF ELEALNDKKH QLEGKLQDIR
501 CRLTTQRQEI ESTNKSREL R IAEITHLQQQ LQESQQMLGR LIPEKQILND
551 QLKQVQQNSL HRDSLVT LKR ALEAKELARQ HLRDQLDEVE KETRSLQEI
601 DIFNNQLKEL REIHNKQQLQ KQKSMEAERL KQKEQERKII ELEKQKEEAQ
651 RRAQERDKQW LEHVQQED EHQPRKLHEEE KLKREESVKK KDGEKKGKQE
701 AQDKLGRLFH QHQEPAPAV QAPWSTAEKG PLTISAQENV KVVYYRALYP
751 FESRSHDEIT IQPGDIVMVK GEWVDESQTG EPGWLGGELK GKTGWFPANY
801 AEKIPENEVP APVKPVT DST SAPAPKLALR ETPAPLAVTS SEPSTTPNNW
851 ADFSSTWPTS TNEKPETDNW DAWAAQPSLT VPSAGQLRQR SAFTPATATG
901 SSPSPVLGQG EKVEGLQAQA LYPWRAKKDN HLNFNKNDVI TVLEQQDMWW
951 FGEVQGQKGW FPKSYVKLIS GPIRKSTSMD SGSSSPASL KRVASPAAP
1001 VVSGEFIAM YTYESSEQGD LTFQQGDVIL VTKKDGDWWT GTVGDKAGVF
1051 PSNYVRLKDS EGS GTAGKTG SLGKKPEIAQ VIASYTATGP EQLTLAPGQL
1101 ILIRKKNPGG W WEGELQARG KKRQIGW FPA NYVKLLSPGT SKITPTEPPK
1151 STALAAVCQV IGM YDYTAQN DDELA FNKGQ IINVLNKEDP DWWKGEVNGQ
1201 VGLFPSNYVK LTTDM DPSQQ *

Figure 9

Whole protein sequence

1 TRGSEGGREE WRRQGRERSL VAP*YGGSRG RIPSGLRDGQ RGGRGWCAGL
51 RLLRPSQRRV SGTDL SLGRQ RGPARR*GVD *QGKSNRTMA QFPTPFGGSL
101 DIWAITVEER AKHDQQFHSL KPISGFITGD QARNFFFQSG LPQPVLAQIW
151 ALADMNNDGR MDQVEFSIAM KLIK LKQGY QLPSALPPVM KQPVAISSA
201 PAFGMGGIAS MPPLTAVAPV PMGSIPVVG M SPTLVSSVPT AAVPPLANGA
251 PPVIQPLPAF AHPAATLPKS SSFSRSGPGS QLNTKLQKAQ SFDVASVPPV
301 AEWAVPQSSR LKYRQLFNH DKTMSGH LTG PQARTILMQS SLPQAQLASI
351 WNLSDIDQDG KLTAE EFILA MHLIDVAMSG QLPPLVPPE YIPPSFRRVR
401 SGSGISVISS TSVDQRLPEE PVLEDEQQQL EK KLPVTFED KKRENFERN
451 LELEKRRQAL LEQQRKEQER LAQLERAEQE RKERERQEQE RKROLELEKQ
501 LEKQRELERQ REEERRKEIE RREAAKRELE RQRQLEWERN RRQELLNQRN
551 KEQEDIVVLK AKKKTLEFEL EALNDKKHQL EGKLQDIRCR LTTQRQEIES
601 TNKSRELRIA EITHLQQQLQ ESQQMLGR LI PEKQILNDQL KQVQQNSLHR
651 DSLVTLKRAL EAKELARQHL RDQLDEVEKE TRSKLQEIDI FNNQLKELRE
701 IHNKQQLQKQ KSMEAERLKQ KEQERKIH EL EKQKEEAQRR AQERDKQWLE
751 HVQQEDEHQR PRKLHEEEKL KREESVKKKD GEEKGKQEAQ DKLGRLFHQH
801 QEPAKPAVQA PWSTA EKGPL TISAQENVKV VYYRALYPFE SRSHDEITI Q
851 PGDIVMVKGE WDESQTGEP GWLGGELK GK TGWF PANYAE KIPENEVPAP
901 VKPVT DSTSA PAPKLALRET PAPLAVTSSE PSTTPNNWAD FSSTWPTSTN
951 EKPETDNWDA WAAQPSLTVP SAGQLRQ RSA FTPATATGSS PSPVLGQGEK
1001 VEG LQAQALY PWRACKDNHL NFNKNDVITV LEQQDMWWFG EVQGQKGWFP
1051 KSYVKLISGP IRKSTSMDSG SSES PASLKR VASPA AKPVV SGEEFIAMYT
1101 YESSEQD LT FQQGDVILVT KKDGDWWTGT VGDKAGVFPS NYVRLKDSEG
1151 SGTAGKTGSL GKKPEIAQVI ASYTATGPEQ LTLAPGQLIL IRKKNPGGWW
1201 EGELQARGKK RQIGWFPANY VKLLSPGTSK ITPTEPPKST ALAAVCQVIG
1251 MYDYTAQNDD ELAFNKGQII NVLNKEDPDW WKGEVNGQVG LFPSNYVKLT
1301 TDMDPSQQ*I ICCPSPPQA* KSSKRPTIPY HCPEG*WEMQ P*SCDFQHDH
1351 LLPSE*KNSL QSSLP HFTLV ACD RNV*VIT CRDRSKNYKN TQSGSFCGF
1401 PSYSN*LSPT FAQVLSIVLK LFLNIYFSFL INKINK*LLC YFGFAKRPTI
1451 KECCMCY*KL FQMSINLR LD VFFHFVQCYQ LNCAVWGFSP LP*KCRGVQY
1501 LCFKDV*NEP N*SEGVCACL CVS AVPC*AC NTSCT*EISS FHGKAITLYD
1551 ALIHLHLILF CTVTL*PHEK ALCVFVRSQI YLVELVFCLG FLILRVCI A*
1601 NQ*TTPLRSL RSTISTVSF* SLLHEVLFQL LFME*PILNK *FS*QERMYR
1651 SLPAINFQCL HFLTRLWNFY RLI*NGAHGP FVC*ICCS*S PVCLLNTSWK
1701 LSIKMPAAHS TENGAGGASS TI*LSS*RLC NAHSPRVLP A LSGGCAGGRV
1751 EVLLLSHGAE SEDLSSSFSC TSVFSRI*M* NI* IYKPAAL TTVIQPFELV
1801 PCIDN*ILHT KVKKKKKK

Figure 9

1 AGAGTGGAGG CGCCAGGGGA GGGAGCGTAG CTTGGTTGCT CCGTAGTACG
51 GCGGCTCGCG AGGAAGAATC CCGAGCGGGC TCCGGGACGG ACAGAGAGGC
101 GGGCGGGGAT GGTGTGCGGG GCTGCGGCTC CTGCGTCCCT CCCAGCGGCG
151 CGTGAGCGGC ACTGATTTGT CCCTGGGGCG GCAGCGCGGA CCCGCCCCGA
201 GATGAGGCGT CGATTAGCAA GGTAAGTA ACAGAACCAT GGCTCAGTTT
251 CCAACACCTT TTGGTGGCAG CTTGGATATC TGGGCCATAA CTGTAGAGGA
301 AAGAGCGAAG CATGATCAGC AGTTCCATAG TTAAAGCCA ATATCTGGAT
351 TCATTACTGG TGATCAAGCT AGAACTTTT TTTTCAATC TGGGTTACCT
401 CAACCTGTTT TAGCACAGAT ATGGGCACTA GCTGACATGA ATAATGATGG
451 AAGAATGGAT CAAGTGGAGT TTTCCATAGC TATGAACTT ATCAAACCTGA
501 AGCTACAAGG ATATCAGCTA CCCTCTGCAC TTCCCCTGT CATGAAACAG
551 CAACCAGTTG CTATTTCTAG CGCACCAGCA TTGGTATGG GAGGTATCGC
601 CAGCATGCCA CCGCTTACAG CTGTTGCTCC AGTGCCAATG GGATCCATTC
651 CAGTTGTTGG AATGTCTCCA ACCCTAGTAT CTTCTGTTCC CACAGCAGCT
701 GTGCCCCCCC TGGCTAACGG GGCTCCCCCT GTTATACAAC CTCTGCCTGC
751 ATTTGCTCAT CCTGCAGCCA CATTGCCAAA GAGTTCTTCC TTTAGTAGAT
801 CTGGTCCAGG GTCACAACTA AACACTAAAT TACAAAAGGC ACAGTCATTT
851 GATGTGGCCA GTGTCCCACC AGTGGCAGAG TGGGCTGTTT CTCAGTCATC
901 AAGACTGAAA TACAGGCAAT TATTCAATAG TCATGACAAA ACTATGAGTG
951 GACACTTAAC AGGTCCCCAA GCAAGAACTA TTCTTATGCA GTCAAGTTTA
1001 CCACAGGCTC AGCTGGCTT AATATGGAAT CTTTCTGACA TTGATCAAGA
1051 TGGAAAACCT ACAGCAGAGG AATTTATCCT GGCAATGCAC CTCATTGATG
1101 TAGCTATGTC TGGCCAAACCA CTGCCACCTG TCCTGCCTCC AGAATACATT
1151 CCACCTTCTT TTAGAAGAGT TCGATCTGGC AGTGGTATAT CTGTCATAAG
1201 CTCAACATCT GTAGATCAGA GGCTACCAGA GGAACCAGTT TTAGAAGATG
1251 AACACAACA ATTAGAAAAG AAATTACCTG TAACGTTTGA AGATAAGAAG
1301 CGGGAGAACT TTGAACGTGG CAACCTGGAA CTGGAGAAAC GAAGGCAAGC
1351 TCTCCTGGAA CAGCAGCGCA AGGAGCAGGA GCGCCTGGCC CAGCTGGAGC
1401 GGGCGGAGCA GGAGAGGAAG GAGCGTAGC GCCAGGAGCA AGAGCGCAAA
1451 AGACAACTGG AACTGGAGAA GCAACTGGAA AAGCAGCGGG AGCTAGAACG
1501 GCAGAGAGAG GAGGAGAGGA GGAAAGAAAT TGAGAGGCGA GAGGCTGCAA
1551 AACGGGAACT TGAAAGGCAA CGACAACTTG AGTGGGAACG GAATCGAAGG
1601 CAAGAACTAC TAAATCAAAG AAACAAAGAA CAAGAGGACA TAGTTGTACT
1651 GAAAGCAAAG AAAAAGACTT TGGAATTTGA ATTAGAAGCT CTAAATGATA
1701 AAAAGCATCA ACTAGAAGGG AAACCTCAAG ATATCAGATG TCGATTGACC
1751 ACCCAAAGGC AAGAAATTGA GAGCACAAC AAATCTAGAG AGTTGAGAAT
1801 TGCCGAAATC ACCCATCTAC AGCAACAATT ACAGGAATCT CAGCAAATGC
1851 TTGGAAGACT TATTCCAGAA AAACAGATAC TCAATGACCA ATTAAACAA
1901 GTTCAGCAGA ACAGTTTGCA CAGAGATTCA CTTGTTACAC TTAAGAGAGC
1951 CTTAGAAGCA AAAGAACTAG CTCGGCAGCA CCTACGAGAC CAACTGGATG
2001 AAGTGGAGAA AGAACTAGA TCAAACTAC AGGAGATTGA TATTTTCAAT
2051 AATCAGCTGA AGGAACTAAG AGAAATACAC AATAAGCAAC AACTCCAGAA

Figure 10

2101 GCAAAAGTCC ATGGAGGCTG AACGACTGAA ACAGAAAGAA CAAGAACGAA
2151 AGATCATAGA ATTAGAAAAA CAAAAAGAAG AAGCCCAAAG ACGAGCTCAG
2201 GAAAGGGACA AGCAGTGGCT GGAGCATGTG CAGCAGGAGG ACGAGCATCA
2251 GAGACCAAGA AAACCTCCACG AAGAGGAAAA ACTGAAAAGG GAGGAGAGTG
2301 TCAAAAAGAA GGATGGCGAG GAAAAAGGCA AACAGGAAGC ACAAGACAAG
2351 CTGGGTCTGC TTTTCCATCA ACACCAAGAA CCAGCTAAGC CAGCTGTCCA
2401 GGCACCCTGG TCCACTGCAG AAAAAGGTCC ACTTACCATT TCTGCACAGG
2451 AAAATGTAAA AGTGGTGTAT TACCGGGCAC TGTACCCCTT TGAATCCAGA
2501 AGCCATGATG AAATCACTAT CCAGCCAGGA GACATAGTCA TGGTGGATGA
2551 AAGCCAAACT GGAGAACCCG GCTGGCTTGG AGGAGAATTA AAAGGAAAGA
2601 CAGGGTGGTT CCCTGCAAAC TATGCAGAGA AAATCCCAGA AAATGAGGTT
2651 CCCGCTCCAG TGAAACCAGT GACTGATTCA ACATCTGCCC CTGCCCCCAA
2701 ACTGGCCTTG CGTGAGACCC CCGCCCCTTT GGCAGTAACC TCTTCAGAGC
2751 CCTCCACGAC CCCTAATAAC TGGGCCGACT TCAGCTCCAC GTGGCCCCACC
2801 AGCACGAATG AGAAACCAGA AACGGATAAC TGGGATGCAT GGGCAGCCCA
2851 GCCCTCTCTC ACCGTTCCAA GTGCCGGCCA GTTAAGGCAG AGGTCCGCCT
2901 TTA CTCCAGC CACGGCCACT GGCTCCTCCC CGTCTCCTGT GCTAGGCCAG
2951 GGTGAAAAGG TGGAGGGGCT ACAAGCTCAA GCCCTATATC CTTGGAGAGC
3001 CAAAAAAGAC AACCCTTAA ATTTTAACAA AAATGATGTC ATCACCCTCC
3051 TGGAACAGCA AGACATGTGG TGGTTTGGAG AAGTTCAAGG TCAGAAGGGT
3101 TGGTTCCCCA AGTCTTACGT GAAACTCATT TCAGGGCCCA TAAGGAAGTC
3151 TACAAGCATG GATTCTGGTT CTTCAGAGAG TCCTGCTAGT CTAAAGCGAG
3201 TAGCCTCTCC AGCAGCCAAG CCGGTCGTTT CGGGAGAAGA ATTTATTGCC
3251 ATGTACACTT ACGAGAGTTC TGAGCAAGGA GATTTAACCT TTCAGCAAGG
3301 GGATGTGATT TTGGTTACCA AGAAAGATGG TGACTGGTGG ACAGGAACAG
3351 TGGGCGACAA GGCCGGAGTC TTCCCTTCTA ACTATGTGAG GCTTAAAGAT
3401 TCAGAGGGCT CTGGAAGTGC TGGGAAAACA GGGAGTTTAG GAAAAAAACC
3451 TGAAATTGCC CAGGTTATTG CCTCATACAC CGCCACCGGC CCCGAGCAGC
3501 TCACTCTCGC CCCTGGTCAG CTGATTTTGA TCCGAAAAAA GAACCCAGGT
3551 GGATGGTGGG AAGGAGAGCT GCAAGCACGT GGGAAAAAGC GCCAGATAGG
3601 CTGGTTCCCA GCTAATTATG TAAAGCTTCT AAGCCCTGGG ACGAGCAAAA
3651 TCACTCCAAC AGAGCCACCT AAGTCAACAG CATTAGCGGC AGTGTGCCAG
3701 GTGATTGGGA TGTACGACTA CACCGCGCAG AATGACGATG AGCTGGCCTT
3751 CAACAAGGGC CAGATCATCA ACGTCTCAA CAAGGAGGAC CCTGACTGGT
3801 GGAAAGGAGA AGTCAATGGA CAAGTGGGGC TCTTCCCATC CAATTATGTG
3851 AAGCTGACCA CAGACATGGA CCAAGCCAG CAATGAATCA TATGTTGTCC
3901 ATCCCCCCT CAGGCTTGAA AGTCCTTTTG TGGCTTTCCT AGTTACTCAA
3951 ATTGACTTTC CCCACCTTT GCACAGGTGC TTTCAATAGT TTTAAAATTA
4001 TTTTAAATA TATATTTTAG CTTTTTAATA AACAAAAATA ATAAATGACT
4051 TCTTTGCTAT TTTGGTTTTG CAAAAAGACC CACTATCAAG GAATGCTGCA
4101 TGTGCTATTA AAAATTGTTT CAAATGTCCA TAAATCTGAG ACTTGATGTA
4151 TTTTTTCATT TTGTCCAGTG TTACCAACTA AATTGTGCAG TTTGGGGCTT
4201 TTCCCCCTTA CCATAGAAGT GCAGAGGAGT TCAGTATCTC TGTTTTAAAG

Figure 10

4251 ACGTATAGAA TGAGCCCAAT TAAAGCGAAG GTGTTTGTGC TTGTTTGTGT
4301 GTATCAGCTG TACCTTGTTG AGCATGTAAT ACATCCTGTA CATAAGAAAT
4351 TAGTTCCTTC CATGGCAAAG CTATTACCTT GTACGATGCT CTAATCATAT
4401 TGCATTTAAT TTTATTTTGC ACAGTGACCT TGAGCCACA TGAGAAAGCA
4451 CTCTGTGTTT TTGTTGCGTC TCAGATTTAT CTGGTTGAGT TGGTGTGTTG
4501 TTTGGGGTTT TTAATTTTGC GTGTTTGCAT AGCATAAAAT CAGTAGACAA
4551 CACCACTGAG GTCGTTACGA TCAACGATAT CCACAGTCTC TTTTGTGCTT
4601 CTGTTACATG AAGTTTTTAT CCAGTTACTT TTCATGGAAT GACCTATTTT
4651 GAACAAGTAA TTTTCTTGAC AAGAAAGAAT GTATAGAAGT CTCCCTGCAA
4701 TTAATTTCCA ATGTTTACAT TTTTAACTA GACTGTGGAA TTTCTACAGA
4751 TTAATATGAA ATGGAGCTCA TGGTCCGTTT GTGTGTTAGA TATGCTGTAG
4801 CTGAAGCCCT GTTTGTCTTT TAAACACTAG TTGGAAGCTC TCAATAAAAA
4851 TGCCTGCTGC TCACAGCACA GAAAATGGGG CAGGGGGAGC CTCAAGCACA
4901 ATCTAGCTGT CCTCCTAAAG ACTCTGTAAT GCTCACTCCC CTCGCGTTCT
4951 CCCGGCGCTG TCGGGAGGCT GTGCTGGTGG TCGTGTAGAG GTCCTTCTCC
5001 TTTACATGG TGCAGAGAGC GAGGACCTCT CCTCCTCGTT CAGTTGCACT
5051 TCAGTATTTT CACGGATATG AATGTAAAAT ATATAAATAT ATAAACCTGC
5101 GGCTTTAACA ACTGTAATAC AACCTTTTGA ATTAGTTCCG TGTATAGATA
5151 ATTAAATTCT TCATACAAAA GTTAAAAAAA AAAAAAAAAA AAAAA

FIGURE 10

Figure 10

Translated Protein Sequence #11

1 MAQFPTPFGG SLDIWAITVE ERAKHDQQFH SLKPISGFIT GDOARNFFQ
 51 SGLPQPVLAAQ IWAADMND GRMDQVEFSI AMKLIKLLQ GYQLPSALPP
 101 VMKQOPVAIS SAPAFGMGGI ASMPPLTAVA PVPMGSIPIV GMSPTLVSSV
 151 PTAAVPPPLAN GAPPVIOPLP AFAHPAATLP KSSFSRSGP GSQNLTKLQK
 201 AQSFDAVSP PVAEWAVPQS SRLKYROLFN SHDKTMSGHL TGPQARTILM
 251 QSSLPQAQLA SIWNLSIDIQ DGKLTAEFI LAMHLIDVAM SGQPLPPVLP
 301 PEYIPPSFRR VRSGSGISVI SSTSDQRLP EEPVLEDEQQ QLEKKLPVTF
 351 EDKKRENFER GNLEKRRQ ALLEQORKEQ ERLAQLERAQ QERKERERQE
 401 QERKQLELE KQLEKQRELE ROREEERRKE IERREAARE LERQRLWE
 451 RNRQELLNQ RNKEQEDIVV LKAKKKTLEF ELEALNDKKH QLEGKLDIR
 501 CRTTQRQEI ESTNKSREL RIAETHLQQQ LQESQOMLGR LIPEKQILND
 551 QLKQVQNSL HRDSLVTLLR ALEAKELARQ HLRDQLDEVE KETRSLQEI
 601 DIFNNQLKEL REIHNNQQLQ KQKSMEARL KQKEQERKII ELEKKEEAQ
 651 RRAQERDKQW LEHVQDEDEH QRPRKLHEEE KLKREESVKK KDGEEKGQKE
 701 AQDKLGRFLH QHQEPAPAV QAPWSTAEG PLTISAQENV KVVYRALYP
 751 FESRSHDEIT IQPGDIVMVD ESQTGEPGWL GGELKGKTGW FPANYAEKIP
 801 ENEVPAPVPV VTDTSAPAP KLALRETPAP LAVTSSEPST TPNWADFSS
 851 TWPTSTNEKP ETDNWDAA QPSLTVPSAG QLRQSAFTP ATATGSSSPSP
 901 VLGGQEKVEG LQAQALYPWR AKKDNHLNFN KNDVITVLEQ QDMWWFGEVQ
 951 GQKGWFPKSY VKLISGPIRK STMSDGSSE SPASLKRVAS PAAKPVVSGE
 1001 EFIAMYTYES SEQGDLTFQO GDVILVTKKD GDWWTGTVD KAGVFPSNYV
 1051 RLKDSGSGT AGKTGSLGKK PEIAQVIASV TATGPEQLT APGQLILRK
 1101 KNPGGWWEGL QARGKKRQI GWFPANYVKL LSPGTSKITP TEPKSTALA
 1151 AVCQVIGMYD YTAQNDELA FNKGQIINV LNKEDPDWWKG EVNGQVGLFP
 1201 SNYVKLT TDM DPSQQ

whole protein sequence:

1 EWRROGRERS LVAP*YGGSR GRIPSGLRDG QRGGRGWCAG LRLLRPSQRR
 51 VSGTDLSLGR QRGPAR*GV D*QGKSNRTM AQFPTPFGGS LDIWAITVEE
 101 RAKHDQQFHS LKPISGFIT GDOARNFFQS GLPQPVLAAQ IWAADMNDG
 151 RMDQVEFSIA MKLIKLLQ GYQLPSALPP VMKQOPVAISS APAFGMGGIA
 201 SMPPLTAVAP VPMGSIPIV GMSPTLVSSV PTAAVPPLANG APPVIOPLP
 251 FAHPAATLP KSSFSRSGP GSQNLTKLQK AQSFDAVSP VAEWAVPQS
 301 RLKYROLFN SHDKTMSGHL TGPQARTILM QSSLPQAQLA SIWNLSIDIQ
 351 GKLTAEFI LAMHLIDVAM SGQPLPPVLP PEYIPPSFRR VRSGSGISVI
 401 STSDQRLP EEPVLEDEQQ QLEKKLPVTF EDKKRENFER GNLEKRRQ
 451 LLEQORKEQ ERLAQLERAQ QERKERERQE QERKQLELE KQLEKQRELE
 501 QREEERRKE IERREAARE LERQRLWE RNRQELLNQ RNKEQEDIVV
 551 KAKKKTLEF ELEALNDKKH QLEGKLDIR CRTTQRQEI ESTNKSREL
 601 RIAETHLQQQ LQESQOMLGR LIPEKQILND QLKQVQNSL HRDSLVTLLR
 651 ALEAKELARQ HLRDQLDEVE KETRSLQEI DIFNNQLKEL REIHNNQQLQ
 701 KQKSMEARL KQKEQERKII ELEKKEEAQ RRAQERDKQW LEHVQDEDEH
 751 QRPRKLHEEE KLKREESVKK KDGEEKGQKE AQDKLGRFLH QHQEPAPAV
 801 QAPWSTAEG PLTISAQENV KVVYRALYP FESRSHDEIT IQPGDIVMVD
 851 ESQTGEPGWL GGELKGKTGW FPANYAEKIP ENEVPAPVPV VTDTSAPAP
 901 KLALRETPAP LAVTSSEPST TPNWADFSS TWPTSTNEKP ETDNWDAA
 951 QPSLTVPSAG QLRQSAFTP ATATGSSSPSP VLGGQEKVEG LQAQALYPWR
 1001 AKKDNHLNFN KNDVITVLEQ QDMWWFGEVQ GQKGWFPKSY VKLISGPIRK
 1051 STMSDGSSE SPASLKRVAS PAAKPVVSGE EFIAMYTYES SEQGDLTFQO
 1101 GDVILVTKKD GDWWTGTVD KAGVFPSNYV RLKDSGSGT AGKTGSLGKK
 1151 PEIAQVIASV TATGPEQLT APGQLILRK KNPGGWWEGL QARGKKRQI
 1201 GWFPANYVKL LSPGTSKITP TEPKSTALA AVCQVIGMYD YTAQNDELA
 1251 FNKGQIINV LNKEDPDWWKG EVNGQVGLFP SNYVKLT TDM DPSQQ
 1301 SPPQA*KSFC GFPSYSN*LS PTFAQVLSIV LKLFLNIYFS FLINKINK*L
 1351 LCYFGFAKRP TKECCMCY*K LKFQMSINLR LDVFFHFVQC YQLNCAVWGF
 1401 SPL*KCRGV QYLCFKDV*N EPN*SEGVCALCLVSAVPC*ACNTSCT*EI
 1451 SSFHGKAITL YDALIILHLI LFCTVTL*PH EKALCVFVRS QIYLVLFVC
 1501 LGFLLIRVCI A*NO*TTPLR SLRSTISTVS F*SLLEHVF QLLFME*PIL
 1551 NK*FS*QERM YRSLPAINFQ CLHFLTRLWN FYRLI*NGAH GPFCV*ICCS
 1601 *SPVCLLNTS WKLSIKMPAA HSTENGAGGA SSTI*LSS*R LCNAHSRVL
 1651 PALSGGCAGG RVEVLLSHG AESEDLSSSF SCTSVFSRI*M*NI*YKPA
 1701 ALTTVIQFFE LVPCIDN*IL HTKVKKKKKK K

Figure 11

1 CGGGGATGGT GTGCGGGGCT GCGGCTCCTG CGTCCCTCCC AGCGGCGCGT
51 GAGCGGCACT GATTTGTCCC TGGGGCGGCA GCGCGGACCC GCCCGGAGAT
101 GAGGCGTCGA TTAGCAAGGT AAAAGTAACA GAACCATGGC TCAGTTTCCA
151 ACACCTTTTG GTGGCAGCCT GGATATCTGG GCCATAACTG TAGAGGAAAG
201 AGCGAAGCAT GATCAGCAGT TCCATAGTTT AAAGCCAATA TCTGGATTCA
251 TTACTGGTGA TCAAGCTAGA AACTTTTTTT TTCAATCTGG GTTACCTCAA
301 CCTGTTTTAG CACAGATATG GGCCTAGCT GACATGAATA ATGATGGAAG
351 AATGGATCAA GTGGAGTTTT CCATAGCTAT GAAACTTATC AAAGTGAAGC
401 TACAAGGATA TCAGCTACCC TCTGCACTTC CCCCTGTCAT GAAACAGCAA
451 CCAGTTGCTA TTTCTAGCGC ACCAGCATT TGGTATGGGAG GTATCGCCAG
501 CATGCCACCG CTTACAGCTG TTGCTCCAGT GCCAATGGGA TCCATTCCAG
551 TTGTTGGAAT GTCTCCAACC CTAGTATCTT CTGTTCCAC AGCAGCTGTG
601 CCCCCCTGG CTAACGGGGC TCCCCCTGTT ATACAACCTC TGCCTGCATT
651 TGCTCATCCT GCAGCCACAT TGCCAAAGAG TTCTTCCTTT AGTAGATCTG
701 GTCCAGGGTC ACAACTAAAC ACTAAATTAC AAAAGGCACA GTCATTTGAT
751 GTGGCCAGTG TCCCACCAGT GGCAGAGTGG GCTGTTCTC AGTCATCAAG
801 ACTGAAATAC AGGCAATTAT TCAATAGTCA TGACAAAAC ATGAGTGGAC
851 ACTTAACAGG TCCCCAAGCA AGAACTATT TATGCAGTC AAGTTTACCA
901 CAGGCTCAGC TGGCTTCAAT ATGGAATCTT TCTGACATTG ATCAAGATGG
951 AAAACTTACA GCAGAGGAAT TTATCCTGGC AATGCACCTC ATTGATGTAG
1001 CTATGTCTGG CCAACCACTG CCACCTGTCC TGCCTCCAGA ATACATTCCA
1051 CTTTCTTTTA GAAGAGTTCG ATCTGGCAGT GGTATATCTG TCATAAGCTC
1101 AACATCTGTA GATCAGAGGC TACCAGAGGA ACCAGTTTTA GAAGATGAAC
1151 AACAACAATT AGAAAAGAAA TTACCTGTAA CGTTTGAAGA TAAGAAGCGG
1201 GAGAACTTTG AACGTGGCAA CCTGGAAGT GAGAAACGAA GGCAAGCTCT
1251 CCTGGAACAG CAGCGCAAGG AGCAGGAGCG CCTGGCCCAG CTGGAGCGGG
1301 CGGAGCAGGA GAGGAAGGAG CGTGAGCGCC AGGAGCAAGA GCGCAAAAGA
1351 CAACTGGAAC TGGAGAAGCA ACTGGAAGAG CAGCGGGAGC TAGAACGGCA
1401 GAGAGAGGAG GAGAGGAGGA AAGAAATTGA GAGGCGAGAG GCTGCAAAAC
1451 GGGAACCTGA AAGGCAACGA CAACTTGAGT GGGAACGGAA TCGAAGGCAA
1501 GAACTACTAA ATCAAAGAAA CAAAGAACAA GAGGACATAG TTGTAAGTAA
1551 AGCAAAGAAA AAGACTTTGG AATTTGAATT AGAAGCTCTA AATGATAAAA
1601 AGCATCAACT AGAAGGGAAA CTTCAAGATA TCAGATGTCT ATTGACCACC
1651 CAAAGGCAAG AAATTGAGAG CACAAACAAA TCTAGAGAGT TGAGAATTGC
1701 CGAAATCACC CATCTACAGC AACAATTACA GGAATCTCAG CAAATGCTTG
1751 GAAGACTTAT TCCAGAAAAA CAGATACTCA ATGACCAATT AAAACAAGTT
1801 CAGCAGAACA GTTTGCACAG AGATTCACTT GTTACACTTA AAAGAGCCTT
1851 AGAAGCAAAA GAACTAGCTC GGCAGCACCT ACGAGACCAA CTGGATGAAG
1901 TGGAGAAAGA AACTAGATCA AACTACAGG AGATTGATAT TTTCAATAAT
1951 CAGCTGAAGG AACTAAGAGA AATACACAAT AAGCAACAAC TCCAGAAGCA
2001 AAAGTCCATG GAGGCTGAAC GACTGAAACA GAAAGAACAA GAACGAAAGA
2051 TCATAGAATT AGAAAAAAAA AAAAAAAAAA

Figure 12

#5 translated Protein sequence:

1 MAQFPTPFGG SLDIWAITVE ERAKHDQQFH SLKPISGFIT GDQARNFFFQ
 51 SGLPQPVLAAQ IWALADMNND GRMDQVEFSI AMKLIKLLKQ GYQLPSALPP
 101 VMKQQPVAIS SAPAFGMGGI ASMPPLTAVA PVPMGSIPVV GMSPTLVSSV
 151 PTAAVPPLAN GAPPVIQPLP AFAHPAATLP KSSSFSGSGP GSQLNTKLQK
 201 AQSFVDASVP PVAEWAVPOS SRLKYRQLFN SHDKTMSGHL TGPQARTILM
 251 QSSLPQAQLA SIWNLSIDIDQ DGKLTAEEFI LAMHLIDVAM SGQPLPPVLP
 301 PEYIPPSFRR VRSGSGISVI SSTSVQRLP EEPVLEDEQQ QLEKKLPVTF
 351 EDKKRENFER GNLELEKRRQ ALLEQQRKEQ ERLAQLERAE QERKERERQE
 401 QERKRQLELE KQLEKQRELE RQREEERRKE IERREAAKRE LERQRQLEWE
 451 RNRQELLNQ RNKEQEDIVV LKAKKKTLEF ELEALNDKKH QLEGKLQDIR
 501 CRLTTQRQEI ESTNKSREL RIAEITHLQQQ LQESQQMLGR LIPEKQILND
 551 QLKQVQQNSL HRDSLVTLLR ALEAKELARQ HLRDQLDEVE KETRSLQEI
 601 DIFNNQLKEL REIHNKQQLQ KQKSMEAERL KQKEQERKII ELEKKKKK

whole sequence

1 RGWCAGLRLR RPSQRRVSGT DLSLGRQGRP ARR*GVD*QG KSNRTMAQFP
 51 TPFGGSLDIW AITVEERAKH DQQFHSCLKPI SGFITGDQAR NFFFQSGLPQ
 101 PVLAQIWALA DMNNDGRMDQ VEFSLAMKLI KKLQGYQLP SALPPVMKQQ
 151 PVAISSAPAF GMGGIASMPP LTAVAPVPMG SIPVVGMSPT LVSSVPTAAV
 201 PPLANGAPPV IQPLPAFAHP AATLPKSSSF SRSGPGSQLN TKLQKAQSFD
 251 VASVPPVAEW AVPQSSRLKY RQLFNSHDKT MSGHLTGPQA RTILMQSSLP
 301 QAQLASIWNL SDIDQDGKLT AEEFILAMHL IDVAMSGQPL PPVLPPEYIP
 351 PSFRRVRSGS GISVISSTSV DQRLPEEPVL EDEQQQLEKK LPVTFEDKKR
 401 ENFERGNLEL EKRRQALLEQ QRKEQERLAQ LERAEQERKE RERQEERKR
 451 QLELEKQLEK QRELERQREE ERRKEIERRE AAKRELERQR QLEWERNRRQ
 501 ELLNQRNKEQ EDIVVLKAKK KTLFELEAL NDKKHQLEGK LQDIRCRLTT
 551 QRQEIESTNK SRELRIAEIT HLQQQLQESQ QMLGRLIPEK QILNDQLKQV
 601 QQNSLHRDSL VTLKRALEAK ELARQHLDQ LDEVEKETRS KLQEIDIFNN
 651 QLKELREIHN KQQLQKQKSM EAERLKQKEQ ERKIIELEKK KKK

Figure 13

1 GACCACCCAA AGGCAAGAAA TTGAGAGCAC AAACAAATCT AGAGAGTTGA
51 GAATTGCCGA AATCACCCAT CTACAGCAAC AATTACAGGA ATCTCAGCAA
101 ATGCTTGGA GACTTATTCC AGAAAAACAG AACTCAATG ACCAATTAAA
151 ACAAGTTCAG CAGAACAGTT TGCACAGAGA TTCATTGTT AACTTAAAA
201 GAGCCTTAGA AGCAAAAGAA CTAGCTCGGC AGCACCTACG AGACCAACTG
251 GATGAAGTGG AGAAAGAAAC TAGATCAAAA CTACAGGAGA TTGATATTTT
301 CAATAATCAG CTGAAGGAAC TAAGAGAAAT ACACAATAAG CAACAACTCC
351 AGAAGCAAAA GTCCATGGAG GCTGAACGAC TGAACAGAA AGAACAAGAA
401 CGAAAGATCA TAGAATTAGA AAAACAAAAA GAAGAAGCCC AAAGACGAGC
451 TCAGGAAAGG GACAAGCAGT GGCTGGAGCA TGTGCAGCAG GAGGACGAGC
501 ATCAGAGACC AAGAAAACCT CACGAAGAGG AAAAACTGAA AAGGGAGGAG
551 AGTGTCAAAA AGAAGGATGG CGAGGAAAAA GGCAACAGG AAGCACAAGA
601 CAAGCTGGGT CGGCTTTTCC ATCAACACCA AGAACCAGCT AAGCCAGCTG
651 TCCAGGCACC CTGGTCCACT GCAGAAAAAG GTCCACTTAC CATTCTGCA
701 CAGGAAAATG TAAAAGTGGT GTATTACCGG GCACTGTACC CCTTTGAATC
751 CAGAAGCCAT GATGAAATCA CTATCCAGCC AGGAGACATA GTCATGGTGG
801 ATGAAAGCCA AACTGGAGAA CCCGGCTGGC TTGGAGGAGA ATAAAAGGA
851 AAGACAGGGT GGTTCCCTGC AAATATGCA GAGAAAATCC CAGAAAATGA
901 GGTTCCCGCT CAGTGAAAC CAGTGACTGA TTCAACATCT GCCCCTGCCC
951 CCAAATGAG CTTGCGTGAG ACCCCCGCCC CTTTGGCAGT AACCTCTTCA
1001 GAGCCCTCCA CGACCCCTAA TAACTGGGCC GACTTCAGCT CCACGTGGCC
1051 CACCAGCACG AATGAGAAAC CAGAAACGGA TAACTGGGAT GCATGGGCAG
1101 CCCAGCCCTC TCTACCGTT CCAAGTGCCG GCCAGTTAAG GCAGAGGTCC
1151 GCCTTTACTC CAGCCACGGC CACTGGCTCC TCCCGTCTC CTGTGCTAGG
1201 CCAGGGTGAA AAGGTGGAGG GGCTACAAGC TCAAGCCCTA TATCCTTGGA
1251 GAGCAAAAAA AGACAACCAC TTAAATTTTA AAAAAATGA TGTCATCACC
1301 GTCCTGGAAC AGCAAGACAT GTGGTGGTTT GGAGAAGTTC AAGGTCAGAA
1351 GGGTTGGTTC CCAAGTCTT ACGTGAAACT CATTTCAGGG CCCATAAGGA
1401 AGTCTACAAG CATGGATTCT GGTTCTTCAG AGAGTCCTGC TAGTCTAAAG
1451 CGAGTAGCCT CTCCAGCAGC CAAGCCGGTC GTTTCGGGAG AAGAAATTGC
1501 CCAGGTTATT GCCTCATACA CCGCCACCGG CCCCAGCAG CTCACTCTCG
1551 CCCCTGGTCA GCTGATTTTG ATCCGAAAAA AGAACCCAGG TGGATGGTGG
1601 GAAGGAGAGC TGCAAGCACG TGGGAAAAAG CGCCAGATAG GCTGGTTCCC
1651 AGCTAATTAT GTAAAGCTTC TAAGCCCTGG GACGAGCAAA ATCACTCCAA
1701 CAGAGCCACC TAAGTCAACA GCATTAGCGG CAGTGTGCCA GGTGATTGGG
1751 ATGTACGACT ACACCGCGCA GAATGACGAT GAGCTGGCCT TCAACAAGGG
1801 CCAGATCATC AACGTCTCA ACAAGGAGGA CCCTGACTGG TGGAAGGAG
1851 AAGTCAATGG ACAAGTGGGG CTCTCCCAT CCAATTATGT GAAGCTGACC
1901 ACAGACATGG ACCCAAGCCA GCAATGAATC ATATGTTGTC CATCCCCC
1951 TCAGGCTTGA AAGTCCTTTT GTGGCTTTCC TAGTTACTCA AATTGACTTT
2001 CCCCCACCTT TGCACAGGTG CTTTCAATAG TTTTAAATTT ATTTTAAAT

Figure 14

2051 ATATATTTTA GCTTTTTAAT AAACAAAATA AATAAATGAC TTCTTTGCTA
2101 TTTTGGTTTT GCAAAAAGAC CCACTATCAA GGAATGCTGC ATGTGCTATT
2151 AAAAATTGTT CCAAATGTCC ATAAATCTGA GACTTGATGT ATTTTTTCAT
2201 TTTGTCCAGT GTTACCAACT AAATTGTGCA GTTTGGGGCT TTTCCCCCTT
2251 ACCATAGAAG TGCAGAGGAG TTCAGTATCT CTGTTTTAAA GACGTATAGA
2301 ATGAGCCCAA TTAAAGCGAA GGTGTTTGTG CTTGTTTGTG TGTATCAGCT
2351 GTACCTTGTT GAGCATGTAA TACATCCTGT ACATAAGAAA TTAGTTCTTT
2401 CCATGGCAAA GCTATTACCT TGTACGATGC TCTAATCATA TTGCATTTAA
2451 TTTTATTTTG CACAGTGACC TTGTAGCCAC ATGAGAAAGC ACTCTGTGTT
2501 TTTGTTCCGT CTCAGATTTA TCTGGTTGAG TTGGTGT TTTT GTTTGGGGTT
2551 TTTAATTTTG CGTGT TTTGCA TAGCATAAAA TCAGTAGACA ACACCACTGA
2601 GGTGTTACG ATCAACGATA TCCACAGTCT CTTTTTAGTC TCTGTTACAT
2651 GAAGTTTTAT TCCAGT TACT TTTCATGGAA TGACCTATTT TGAACAAGTA
2701 ATTTTCTTGA CAAGAAAGAA TGTATAGAAG TCTCCCTGCA ATTAATTTC
2751 AATGTTTACA TTTTTTAACT AGACTGTGGA ATTTCTACAG ATTAATATGA
2801 AATGGAGCTC ATGGTCCGTT TGTGTGTTAG ATATGCTGTA GCTGAAGCCC
2851 TGTTTGTCTT TTAAACACTA GTTGGAAGCT CTCAATAAAA ATGCCTGCTG
2901 CTCACAGCAC AGAAAATGGG GCAGGGGGAG CCTCAAGCAC AATCTAGCTG
2951 TCCTCCTAAA GACTCTGTAA TGCTCACTCC CCTCGCGTTC TCCCGGCGCT
3001 GTCGGGAGGC TGTGCTGGTG GTCGTGTAAG GTCCTTCTCC TTTCACATGG
3051 TGCAGAGAGC GAGGACCTCT CCTCCTCGTT CAGTTGCACT TCAGTATTTT
3101 CACGGATATG AATGTAAAAAT ATATAAATAT ATAAACCTGC GGCTTTAACA
3151 ACTGTAATAC AACCTTTTGA ATTAGTTCCG TGTATAGATA ATTAAATTCT
3201 TCATACAAAA GTTAAAAAAA AAAAAAAAAA A

Figure 14

#9 translated protein sequence:

1 TTQRQEIEST NKSRELRIAE ITHLQQQLQE SQQMLGRLIP EKQILNDQLK
51 QVQQNSLHRD SLVTLKRALE AKELARQHLR DQLDEVEKET RSKLQEIDIF
101 NNQLKELREI HNKQQQLQKQK SMEAERLKQK EQERKIIIELE KQKEEAQRRRA
151 QERDKQWLEH VQQEDQHQR RKLHEEEKLK REESVKKKDG EEKGKQEAQD
201 KLGRLFHQHQ EAPKPAVQAP WSTAEGPLT ISAQENVKVV YYRALYPFES
251 RSHDEITIQP GDIVMVDESQ TGEPGWLGGG LKGKTGWFP NYAEKIPENE
301 VPAPVKPVT DSTSAPAPKLA LRETPAPLAV TSSEPSTTPN NWADFSSTWP
351 TSTNEKPETD NWDAWAAQPS LTVPSAGQLR QRSFTPTATA TGSSPSPVLG
401 QGEKVEGLQA QALYPWRACK DNHLNFNKND VITVLEQQDM WWFGEVQGQK
451 GWFPKSYVKL ISGPIRKSTS MDSGSSSPA SLKRVASPAA KPVVSGEEIA
501 QVIASYTATG PEQLTLAPGQ LILIRKKNPG GWWEGELQAR GKRRQIGWFP
551 ANYVKLLSPG TSKITPTEPP KSTALAAVCQ VIGMYDYTAQ NDELAFNKG
601 QIINVLNKED PDWWKGEVNG QVGLFPSNYV KLTTDMDPSQ Q*

Whole protein sequence

1 TTQRQEIEST NKSRELRIAE ITHLQQQLQE SQQMLGRLIP EKQILNDQLK
51 QVQQNSLHRD SLVTLKRALE AKELARQHLR DQLDEVEKET RSKLQEIDIF
101 NNQLKELREI HNKQQQLQKQK SMEAERLKQK EQERKIIIELE KQKEEAQRRRA
151 QERDKQWLEH VQQEDQHQR RKLHEEEKLK REESVKKKDG EEKGKQEAQD
201 KLGRLFHQHQ EAPKPAVQAP WSTAEGPLT ISAQENVKVV YYRALYPFES
251 RSHDEITIQP GDIVMVDESQ TGEPGWLGGG LKGKTGWFP NYAEKIPENE
301 VPAPVKPVT DSTSAPAPKLA LRETPAPLAV TSSEPSTTPN NWADFSSTWP
351 TSTNEKPETD NWDAWAAQPS LTVPSAGQLR QRSFTPTATA TGSSPSPVLG
401 QGEKVEGLQA QALYPWRACK DNHLNFNKND VITVLEQQDM WWFGEVQGQK
451 GWFPKSYVKL ISGPIRKSTS MDSGSSSPA SLKRVASPAA KPVVSGEEIA
501 QVIASYTATG PEQLTLAPGQ LILIRKKNPG GWWEGELQAR GKRRQIGWFP
551 ANYVKLLSPG TSKITPTEPP KSTALAAVCQ VIGMYDYTAQ NDELAFNKG
601 QIINVLNKED PDWWKGEVNG QVGLFPSNYV KLTTDMDPSQ Q*HCCPSPP
651 QA*KSFCGFP SYSN*LSPTF AQVLSIVLKL FLNIYFSFLI NKINK*LLCY
701 FGFAKRPTIK ECCMCY*KLF QMSINLRDLV FFHFVQCYQL NCAVWGFSP
751 P*KCRGVQYL CFKDV*NEPN *SEGVACLC VSAVPC*ACN TSCT*EISSF
801 HGKAITLYDA LILHLILFC TVTL*PHEKA LCVFVRSQIY LVELVFCLGF
851 LILRVCIAN Q*TTPLRSLR STISTVSF*S LLHEVLFLQLL FME*PILNK*
901 FS*QERMYRS LPAINFQLH FLTRLWNFYR LI*NGAHGPF VC*ICCS*SP
951 VCLLNTSWKL SIKMPAAHST ENGAGGASST I*LSS*RLCN AHSRVLPLAL
1001 SGGCAGGRVR SFSFHMVQRA RTSPPRSVAL QYFHGYECKI YKYNLRL*Q
1051 L*YNLLN*FR V*IKFFIQK LKKKKK

Figure 15

09/720934 0004



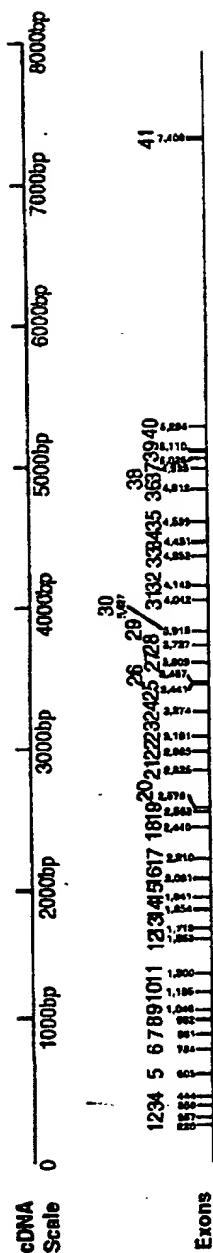
Mouse E9
Tissue

Embryo day 9

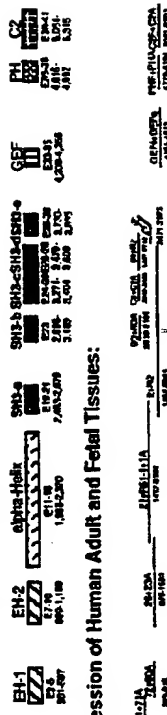
Figure 16

Summary of Studies on ITS (Intersectin) AKA SH3P17

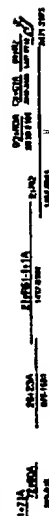
I. Gene Sequence:



7. Protein Domains vs. Nucleotide sequence:



III. Gene Expression of Human Adult and Fetal Tissues:



Northern Blots

15Kb	(F)	-	-	-	-	+	(AB)	-
9.0Kb	(F)	-	-	-	-	-	-	-
5.4Kb	+	+	+	+	+	+	+	+
4.5Kb	(F)	+	(F)	+	+	+	+	+
2.0Kb	+	+	+	+	+	+	+	+

* Human ITS (Intersectin), AKA SH3P17 is ubiquitously expressed with extensive alternative splicing generating tissue and developmental stage-specific expression.

IV. Gene Expression with Antibodies to SH3-e:

* Gene expression is specific to subpopulation of neurons during CNS morphogenesis and in fetal liver, suggesting possible roles for this gene in hematopoiesis, possibly leukemia and platelet formation as well as in brain formation.

B= band seen only in adult and fetal brain
AB= band seen only in adult brain
FB= band seen only in fetal brain
FL= band seen only in fetal liver

Figure 17

2320-1-001 PCT

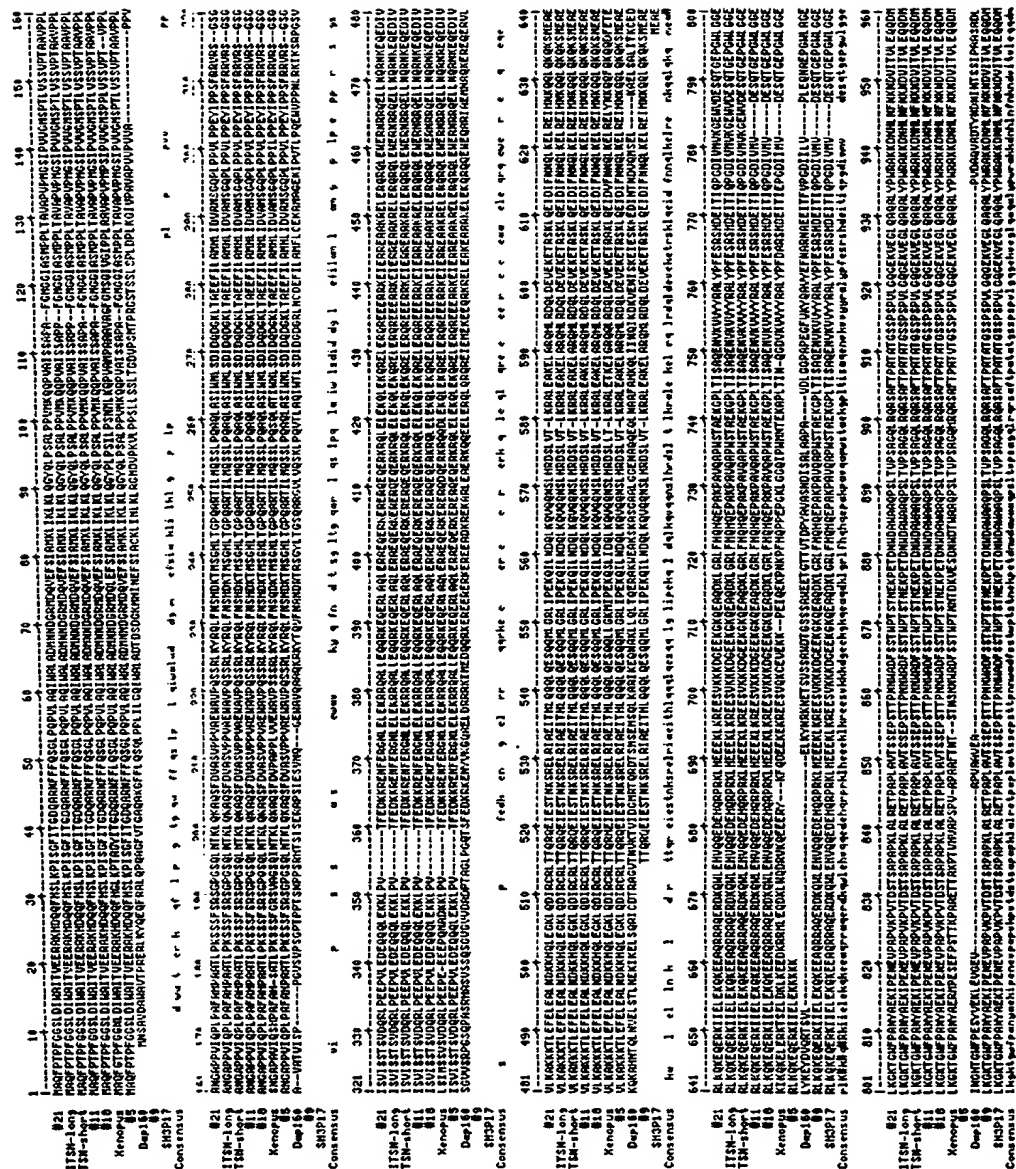


Figure 18

2320-1-001 PCT
(Sheet 30 of 30)